PCT/GB 0 0 / 0 3 5 6 8

The Patent Office

INVESTOR IN PEOPLE

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

GB0003568

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

4

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Andrew Gerrey

Patents Form 1/77

The Patent Office

Request for grant of a patent

9921938.8

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1 Your reference

LPB/P32059

2 Patent application number

Full name, address and postcode of the applicant

The University of York
Heslington Hall
YORK
YO10 5DD

Patents ADP number

State of incorporation

UK 05978549001.

4 Title of the invention

Target for Antiviral Therapy

5 Name of agent

HARRISON GODDARD FOOTE

Address for service

Belmont House
20 Wood Lane
LEEDS
LS6 2AE

Patents ADP number

14571001

6 Priority applications | Country | Priority App No | Date of Filing

Patents Form 1/77

7	Parent application (eg Divisional)	Earlier Application No	Date of Filing
8	Statement of Inventorship Needed?	YES	
9	Number of sheets for any of the following (not counting copies of same document)		
	Continuation sheets of this form		
	Description	**25**	
	Claims		
	Abstract		
	Drawings	**13** +13	
10	Number of other documents attached		
	Priority documents		
	Translations of priority documents		
	P7/77		
	P9/77		
	P10/77		
	Other documents		
11	I/We request the grant of a patent on the basis of this application.		
	Signature Lisa Brown		**16 September 1999**
12	Name and daytime telephone number of person to contact in the United Kingdom	**Dr Lisa Brown** **0113 225 8350**	

Target for Antiviral Therapy

The present invention provides a crystallised module of a nuclear phosphoprotein and an assay and method of determining interactions with human papillomavirus E2 for use in drug design, for use particularly but not exclusively, in designing antiviral agents with potential use in treating warts, proliferative skin lesions and carcinoma of the cervix.

Background to the Invention

Human papillomaviruses (HPVs) cause warts and proliferative lesions in skin and other epithelia. In a minority of HPV types ("high risk", which include HPVs 16, 18, 31, 33, 45 and 56), further transformation of the wart lesions can produce tumours, most notably carcinoma of the cervix[1]. HPVs have evolved a sophisticated system of control, mediated by protein:DNA and protein:protein interactions, that involves both cellular and viral proteins. The 45 kDalton nuclear phosphoprotein, E2, has two central roles in this control. It acts as the principal virally encoded transcription factor and, in association with the viral E1 protein, it creates the molecular complex at the origin of the viral DNA replication[2].

E2 has three distinct modules. The N-terminal module (E2NT) of about 200 amino acids is responsible for interactions with viral and host cell transcription factors. It is followed by a flexible, proline-rich, linker module and a C-terminal module (E2CT), each of about 100 amino acids [3] (Fig. 1a). The E2CT binds as a homodimer to DNA sites with a consensus sequence of $ACCGN_4CGGT$ [4]. In most HPVs a long upstream regulatory region (URR) precedes the viral genes and contains four spatially conserved E2 binding sites: three sites proximal to the transcription start site (p97 in

The dimer of E2CT serves to anchor E2 protein to its recognition sites on the DNA; the function of the E2NT is to bind and localise at least three cellular transcription

factors, Sp1, TFIIB and AMF-1, to the transcription initiation complex. In addition, E2 interacts with another viral protein, E1, which has ATPase and helicase activities. E1 itself binds to the viral origin of replication which consists of about 100 bp and is surrounded by the three E2-binding sites, proximal to the transcription start. The E2:E1 interaction greatly increases the rate of HPV genome replication[2,5,6], Fig. 1a. An intact E2 is essential for the normal productive (wart) life cycle of HPV, however during malignant progression HPV DNA is integrated into the host cell genome, which usually results in disruption of the E2/E1 ORFs and loss of E2 protein, in turn leading to dysregulated expression of the viral oncogenes E6 and E7[7].

Consistent with its role as a transcription regulator, E2 has been shown to direct the formation of loops in DNA containing E2 binding sites[8]. The loops were only formed with intact E2, and not with the E2CT alone. The E2 binding sites did not function independently and their co-operative effect was mediated by full length E2, leading the authors to suggest that there were specific interactions mediated by E2 that bridged across the set of DNA binding sites through its N-terminal. A similar DNA loop structure could also be achieved with Sp1, a cellular transcription factor, which forms a complex with distally bound E2 [9]; Sp1/E2 interactions are critical for transcription activation in BPV[10].

Eighty six known E2 proteins from different species and different human subtypes[11] are highly conserved, with sequence identities typically of 35% in the N and C-terminal modules (Fig. 1b). The crystal structure of the E2CT has been determined both alone and in complex with cognate DNA[12-14]. The module is a dimer with a barrel fold, and induces substantial bending (42-44°) of the DNA from its B-form double helix[14].

The structure of the proteolytic fragment of HPV18 E2NT, missing 65 N-terminal residues, was recently reported at 2.1 Å spacing[15]. This allowed some analysis of mutational effects on function, although the missing 65 amino acids contain residues which are essential for the transcriptional and replication activities of the protein.

We report herein the structure of the complete E2NT determined by X-ray analysis at 1.9 Å. We have found that it is an L-shaped molecule with the residues vital for transcriptional and replication activities of the protein lying on opposite sides of the N-terminal domain. Surprisingly, our results show that the surface, vital for transcription activation, is in fact involved in association of two E2NT's into a dimer. We suggest that dimerisation of E2NT plays an important and key role in induction of DNA loop formation, the mechanism by which distally bound transcription factors would be brought close to the site of transcription initiation. More importantly, our results raise the possibility that dimer formation serves as a molecular switch between early gene expression and viral genome replication during HPV infection.

Statement of the Invention

According to a first aspect of the invention there is provided a crystallised molecular complex of an E2 N-terminal module (E2NT) dimer protein or homologue thereof, comprising residues vital for transcriptional and replicational activities of said protein lying on opposite sides of an N-terminal domain, for use in rationalised drug design.

Preferably the E2NT dimer protein is substantially as depicted in any of Figures 2c and/or 3a-d.

According to a second aspect of the invention there is provided an *in vitro* method for identifying and/or selecting a candidate therapeutic agent, the method comprising determining interaction of a E2 N-terminal module (E2NT) dimer in a sample by contacting said sample with said candidate therapeutic agent and measuring DNA loop formation.

Preferably, the candidate therapeutic agent interferes or blocks interactions of E2NT so as to interfere or block viral and/or cellular transcription factors.

According to a third aspect of the invention there is provided use of an E2NT dimer in the preparation of a medicament for use in treating warts, proliferative skin lesions and/or cervical cancer.

According to a fourth aspect of the invention there is provided a method of monitoring the efficacy of an antiviral therapy in a patient receiving a medicament for the treatment of warts, proliferative skin lesions and/or cervical cancer comprising taking a sample from said patient and measuring E2NT interactions and/or DNA loop formation.

Thus it will be appreciated that a patient can be monitored at the start of therapy to test its effectiveness. Alternatively, a patient can be monitored once a therapy has been established so as to monitor its efficacy with a view to altering a therapy if found to be unsatisfactory.

The human papillomavirus E2 protein controls the primary transcription and replication of the viral genome. Both activities are governed by a ~200 amino acid N-terminal module (E2NT) which is connected to a DNA binding C-terminal module by a flexible linker. The crystal structure of the E2NT module from high-risk type 16 human papillomavirus reveals an L-shaped molecule with two closely packed domains, each with a novel fold. It forms a dimer in the crystal and in solution. The dimer structure is important in the interactions of E2NT with viral and cellular transcription factors and is the key to induction of DNA loops by E2. These loops may serve to target distal DNA-binding transcription factors to the region proximal to the start of transcription. The structure has implications for antiviral drug design and cervical cancer therapy.

Detailed Description of the Invention

The invention will now be described by way of example only with reference to the following Figures and Tables wherein:

Table 1 illustrates X-ray data and phasing statistics;

Table 2 illustrates refinement and model correlation;

Figure 1a represents functional assignments of HPV 16 E2 protein;

Figure 1b, illustrates sequence alignment of E2NT modules from a subset of HPV types;

Figure 2a illustrates a stereo view of electron density with a final model at the dimer interface of the E2NT module, viewed down the crystallographic two-fold axis;

Figure 2b represents a stereo ribbon diagram of the E2NT module;

Figure 2c represents the E2NT dimer;

Figure 3a illustrates a schematic view of URR;

Figure 3b illustrates a schematic view of loop formation induced by binding of E2 proteins to two cognate sites;

Figure 3c illustrates a model of E2 dimer formation;

Figure 4a illustrates the distribution of conserved residues on the E2NT monomer;

Figure 4b illustrates a first cluster of conserved residues on the E2NT monomer;

Figure 4c illustrates a second cluster of conserved residues on the E2NT monomer; and

Figure 4d illustrates conserved residues Gln12 and Glu39.

With reference to Figure 1a and functional assignments of E2. There is shown in a schematic view of NT, linker and CT modules of E2 indicating known functions of each module. Amino acid numbers which delimit the modules corresponds to E2 from HPV16. In Figure 1b, there is shown the sequence alignment of the E2NT modules from a subset of HPV types (HPV16, HPV18, HPV11 and HPV2a) and one BPV type. Shaded blocks above the alignment indicate the experimentally determined secondary structure. Shaded blocks below the sequences indicate the minimal peptide sequences involved in protein:protein interactions, suggested by mutation studies. Residues with more than 90% identity among 86 PV types are coloured: red for internal structural residues, green for residues within the fulcrum region, blue for surface residues.

With reference to the structural features of E2, in Figure 2a there is shown a stereo view of the electron density with the final model, at the dimer interface of the E2NT module, viewed down the crystallographic two-fold axis. The likelihood weighted map is contoured at the 1.5 σ level. Ribbons of two independent monomers are coloured blue and yellow. Side chains of ARG37 and Ile73 which are known to be critical for transactivation [4,31] , are shown in dark green; side chain of other residues at the dimer interface are shown in light green. Oxygen atoms are in red, nitrogen in blue, water molecules are shown as orange spheres and hydrogen bonds as dashed sticks. In Figure 2b, there is shown a stereo ribbon diagram of the E2NT module. The N1 domain is shown in aquamarine and the N2 domain in pink, with the fulcrum in green. In Figure 2c, there is shown the dimer of E2NT, showing the extent of the interface between the two subunits. The view is as in Figure 2a but rotated clockwise

by 90°. Side chains of Gln12 and Glu39 which are critical for interactions with E1 [31-33,37] are shown in magenta. Side chains of residues at the dimer interface are coloured as per Figure 2a.

With reference to Figures 3a-d there is shown loop formation in the URR of HPV16. In Figure 3a, there is shown a schematic view of the URR. The four E2-binding sites are represented by boxes. Numbers in italics indicate distances between individual sites upstream of the p97 promoter. Two possible E2 configurations, with separate or dimeric E2NT modules are shown. In Figure 3b, there is shown a schematic view of loop formation induced by binding of E2 proteins to two cognate sites, based on the experiments reported by Knight *et al*[8]. In Figure 3d, there is shown the possible DNA loops within the URR as depicted in Figure 3b. In Figure 3c, there is shown a model of the formation of E2 dimers, showing interactions between both the C-terminal and E2NT modules. The C-terminal dimer, with its bound DNA, is based on the crystal structure of this module[12]. The E2NT dimer is proposed from the present work. The relative orientation and position of the E2NT and C-terminal modules is purely schematic.

With reference to Figures 4a-d there is shown functionally important residues. In Figure 4a, there is shown the distribution of conserved residues on the E2NT monomer. In Figures 4b and 4c there is shown the two clusters of conserved residues in the fulcrum of E2NT. In Figure 4d, there is shown conserved residues Gln12 and Glu39. Bonds in ball-and stick models are coloured aquamarine (N1 domain), pink (N2 domain) and green (fulcrum). Hydrogen bonds are shown as dashed lines, water molecules as orange spheres, oxygen atoms are in red, nitrogen atoms in blue and sulphur atoms in yellow.

There is convincing evidence that the E2 protein has an extended structure, is flexible

problem is the extended flexible linker module, with around 100 residues. E2NT

proved difficult to crystallise, and a number of different constructs were made and overexpressed before crystallisation with residues 1 to 201 was achieved, but even this construct possessed limited stability. The protein had to be crystallised within 2-3 days of purification; crystals grew within about 48 hours but only retained useful diffraction quality for a further 2-3 days. This necessitated that crystals be rapidly vitrified in cryoprotectant buffer and stored for use as soon as detector time became available[16].

Crystals of E2NT belong to the space group $P3_121$ with unit-cell dimensions a=b=54.3 Å, c=155.5 Å. The structure was determined using two heavy atom derivatives and refined with data extending to 1.9 Å spacing (Fig. 2a). The main chain is well defined throughout with the exception of residues 125 and 126 which are in an exposed loop and are mobile. There was density for the last residue of the His-tag at the N-terminus, but none for the remainder of this entity. All amino acids lie in the allowed regions of the Ramachandran (ϕ,ψ) plot[17] with 92.4% in most favoured regions[18].

The transactivation module is composed of two domains, N1 and N2, arranged so as to give it an overall L-shaped appearance. Analysis of the PDB[19]using DALI[20]shows that both have unique organisation of their secondary structures. Domain N1, which forms the N-terminus of the intact E2, is composed of residues 1 to 92, which fold into three long α-helices, Figure 2 (b,c). There is a tight loop between α1 and α2 and a more extended one between α2 and α3. The three helices pack antiparallel to one another in the form of a twisted plane, with angles of about 20° and 25° between the pairs of consecutive helices. DALI indicated a maximum Z-score of 5.7, that could suggest a significant correlation, for colicin 1a, a membrane protein which contains three 80 Å long α-helices arranged more or less coplanar[21]. This is the only other known protein that contains a true domain made up of such a packing of three helices. In addition there were 42 other structures which gave Z-scores above 4.0, most of which were four helix bundles, such as bacterioferritin[22]. However, in these only two of the three N1 helices superimposed simultaneously on two, not always

adjacent, bundle helices as a result of a more planar arrangement of helices within N1. The indications are that the similarities observed reflect the optimum stacking angle of antiparallel helices against one another rather than suggesting a common ancestor for the evolution of these molecules.

Domain N2 is made up of residues 110 to 201 and is composed almost entirely of antiparallel β structure, with only one short helical segment from residues 171 to 178, Figure 2 (b,c). The secondary structure has two short three and four stranded antiparallel β pleated sheets interconnected by two stranded β ribbons. For this domain DALI failed to identify any significant homologies to known structures, with a highest Z-score of only 2.1. From the analysis of Harris and Botchan[15] and the present study, the N2 fold appears to be novel.

The structure between the N1 and N2 domains (residues 93 to 109) contains two consecutive single turns of helical structure, resulting in a compact and tight turn. It packs closely against elements of both domains and is not a truly independent structural domain. Rather it forms a fulcrum in the L-shape formed by N1 and N2 where it could act as a hinge, allowing the two domains to change their relative conformation in a specific way. Several of the interactions between adjacent regions of chain in the fulcrum are mediated indirectly through H-bonds involving water molecules, suggesting the possibility of flexibility.

One of the most striking features of the crystal structure is the association of two E2NT monomers into a tight dimer. The two E2NT monomers pack around the crystallographic 2-fold axis, as shown in Figure 2a. The dimer interface is formed mostly by amino acids from helices α2 and α3 of the N1 domain and by residues 142-144 from the N2 domain. The total buried surface area between the two E2NT is

^{2} ^{2} [12] which are

In the E2NT dimer interface, each subunit contributes a cluster of seven equivalent residues, invariant or conserved in the 86 known sequences of E2[11], with many direct and water-mediated hydrogen bonds and rather few non-polar contacts, Fig. 2. Analysis of the dimer forming surfaces shows that all the direct hydrogen bonds between monomers are made through these seven amino acids. For the invariant Arg37, all possible side-chain hydrogen bonds are made and all are well defined, Figure 2. Three of them are across the dimer interface. One hydrogen bond is critical, from NH2 to the main chain carbonyl oxygen of Leu77. A second hydrogen bond from NH2 is to OG1 of Thr81; in five out of 86 sequences this residue is glutamine, and modelling shows a hydrogen bond is possible to the NE of Arg37. The NH1 of Arg71 H-bonds to the OE1 of residue 80, which is Glu or Gln in all but six variants. At the NE of Arg37 there is an ideal H-bond to water that itself makes another strong H-bond across the dimer interface to the main-chain carbonyl oxygen of residue 142. The role of the invariant Ile73 is the filling of the intersubunit non-polar volume made up of the aliphatic parts of Arg37, Gln76 and of Leu77 - in this case from both monomers. The Leu77 is in a few sequences substituted by valine or isoleucine and in 9 out of 86 known sequences by methionine. Inspection of the structure shows that Leu77 is partially exposed to the solvent and therefore different hydrophobic side chains could be easily accommodated at this site. Another important non-polar side chain is Ala69. Its side chain methyl packs into the surface of the other monomer at van de Waals distance from the main chain of residue 142. The only observed mutation of Ala69 is to Gly, and is easily accommodated. Gln76 is conserved or has homologous substitutions in about 2/3 of E2 sequences; in about 1/4 of the sequences there is methionine or valine at this position[11]. Although hydrophobic substitutions of Gln76 would disrupt the hydrogen bonding to Glu80 across the dimer interface, and to Arg37 from the same subunit, the hydrophobic side chain at residue 76 could instead make a compensating hydrophobic interaction with the adjacent intersubunit hydrophobic pocket formed by Ile73 and Leu77.

Modelling of the amino acid variations in the 86 known papillomavirus E2 proteins into the other contacts at the dimer interface shows that they generally can be accommodated (data not shown). The consistency of the hydrogen bonds and van de

Waals contacts at the monomer-monomer interface in the various sequences suggests therefore that the E2NT dimer interactions are potentially present in all papillomaviruses.

The first experimental evidence for the E2NT dimerisation in the presence of DNA with multiple E2-binding sites was provided by Knight *et al* in 1991[8]. Their studies showed that intact E2 led to the formation of DNA loops on templates with widely separated E2 binding sites, while a truncated E2, containing the DNA-binding E2CT but missing the N-terminal 161 residues, did not. Such dimerisation is further supported by the observed synergistic transcription activation by a complex of two DNA-bound E2 dimers[25].

To analyse the functional behaviour of the E2NT dimers further, we measured the dissociation constant by sedimentation equilibrium using analytical ultracentrifugation of recombinant E2NT protein containing the 201 N-terminal amino acids. A value of $K_d = 8.1 \pm 4 \times 10^{-6}$ M was obtained, indicating medium-strength association. The micromolar range of the E2NT dimer K_d is certainly physiologically significant, and compares well with values for other transcription factors which have relatively low dissociation constants, often with the K_d values between 1 μM and 20 μM [26,27]. *In vivo*, the interaction could be enhanced when the two E2NT modules are placed in close proximity. Indeed, E2CT forms dimers which bind to the multiple DNA-binding sites located within the URR of viral DNA with K_d of protein:DNA interactions usually in the nanomolar range[28]. Consequently, the local concentration of E2NT, bound to the E2CT via the non-conserved, flexible ~80 amino-acid linker, is effectively increased.

E2NT dimer interactions, as seen in the crystal structure, could form either between modules which are already part of a single E2 dimer, formed as a result of E2CT between two preformed E2 dimers located on different E2 binding sites (Fig. 5b). The results of the electron microscopy suggest that the latter dimerisation does

occur[8]. Although no direct experimental evidence exists for the former dimerisation, it does also seem possible due to the flexibility of the linker connecting the two modules. We propose that E2 molecules may initially keep their N-terminal modules within their internal dimers, but swap N-terminal modules and cross link to E2 molecules bound to distant DNA binding sites to form active loop structures during transcriptional activation and / or HPV DNA replication (Figure 3d). As discussed below, the effects of mutations on transcriptional transactivation can be explained in terms of the dimer being an essential element in this process.

E2 is a regulator of both transcription and viral DNA replication and thus interacts with other viral and host macromolecules in the infected cell. Indication of the possible importance of individual residues in the function comes firstly from the structure, secondly from the extensive set of sequences of the papillomaviral E2's and thirdly from mutagenesis studies on the individual proteins. In the following we make a primary attempt to map the molecule's function onto its structure.
The pattern of amino acid conservation for the 86 available papilloma sequences[11] has been analysed using the GCG program suite[29]. The sequences exhibit striking variation, characteristic of some virus families. However, 33 of the total 201 residues in the E2NT construct were totally or highly conserved. Fig. 4a illustrates the distribution of these 33 residues in the dimer. These were categorised into two sets: those with an essentially structural role and those exposed on the surface with a potential for intermolecular interactions. Thirteen residues (Fig. 1b) are buried or play a purely structural role within the monomer, they are not expected to be of functional importance and will not be discussed here.

A further 12 of these 33 residues stand out as having a structural role in the interface of the N1 and N2 domains. They form three clusters, the first making direct interactions between the two domains (Ile82, Glu90, Trp92, Lys112, Tyr138, Val145) and two separate sets of interactions, one from N2 (Pro106, Lys111, Phe168, Trp134) and the other from N1 (Trp33, Leu94) to the structure connecting them, referred to here as a fulcrum. The first two clusters are shown in Figure 4 b, c and it

can be seen that Lys111 and Lys112 play key roles. Their side chains point in opposite directions to one another and their terminal amino groups are involved in near ideal patterns of hydrogen bonds. The flat surfaces of their extended side chains stack against Trp134 and Trp92, respectively. This clustering of invariant residues at the interface indicates a functional importance for the relative orientation of N1 and N2. The fulcrum could indeed provide a flexible pivot between the two domains, but there is no direct evidence for this as yet. Finally, while the side chain of Glu90 is held tightly in place by two H-bonds and could have a structural role, its OE2 atom is exposed on the surface and is surrounded by near invariant side-chains, which may thus play a part in interactions with other molecules.

Of the remaining eight conserved residues, mutational substitutions of Glu20, Glu100 and Asp122 [30-33] had moderate effects on the transactivation and replication properties of E2, which depended on a particular viral strain. Glu20 lies on the top surface of N1. Asp122 lies far away on the distal surface of N2. Glu100 is completely exposed and points into the solvent at the junction of the L between the N1 and N2 domains. The functional role of these amino acids has yet to be clarified.

Three conserved amino acids (Arg37, Glu39 and Ile73) have been subjected to point mutation and the effects on the two principal functions of E2, i.e. transactivation and HPV DNA replication have been assessed (reviewed in[4],also [31,34,35]). Together with the remaining two conserved amino acids, Gln12 and Ala69, these residues form two functionally important surfaces (see below).

Finally, a number of the mutational results (reviewed in [4], also [31,34,35]) correspond to residues that can be assigned to structural roles. Substitution of these residues will lead to substantial conformational changes and a probable inability to fold correctly. This is particularly true for some of the deletion mutants involving the core of the

mutants in the future.

The induction of DNA loops by E2NT dimerisation could be important for the construction of the active transcription bubble by targeting DNA-binding transcription factors, bound at distal sites, to the region proximal to the start of transcription (reviewed in [36]). In support of this, residues Arg37, Ile73 and Gln76 map onto the surface of E2NT involved in dimer formation, and mutations result in considerable disruption of transactivation, while having little effect on replication, [4,15,31]. The structure also shows that Ala69 which points its side chain methyl across the dimer interface, is also critical for transactivation. Mutational substitutions to amino acids with longer side chains should have a knock out effect on E2NT dimer formation and consequently on transactivation.

The sites of association with cellular transcription factors AMF-1 (residues 74-134) and TFIIB (134-216) were previously mapped onto the E2NT module (Figure 1) using a series of deletion mutants as well as point mutations[34,35]. These sites were mutually exclusive. In the structure, residues 74-134 include the fulcrum, while residues 134-216 correspond to domain N2. Further biochemical and structural studies can now be planned to characterise these interactions in more detail.

Replication of the viral genome is initiated by binding of another viral protein, E1, to the origin of DNA replication[4] which is itself flanked by two E2 binding sites, Fig. 3a. While the function of E2CT dimers is to bind specifically to the DNA sites, E2NT interaction with E1 enhances the binding of E1 to this region. Mutational substitutions of Glu39 generally retained transcriptional activation while DNA replication was substantially reduced[31-33,37]. In the structure, the conserved Glu39 makes every possible hydrogen bond by its side chain carboxyl oxygens, Fig. 4d. One hydrogen bond is to NE2 of Gln12, which is absolutely conserved in all known sequences of E2. The other three hydrogen bonds are to the water molecules which are part of an intimate net of well-defined water molecules surrounding Glu39 and mediating its interactions with adjacent residues. Interestingly, a number of these protein interactions with water molecules are conserved as they are made to the

protein backbone, including carbonyl oxygens of Gln12, Met36 and Lys68. While mutation of Gln12 in BPV1 only slightly affected both transactivation and replication, it substantially reduced cooperative origin binding[30,32]. The close positioning of Gln12 and Glu39 in the three-dimensional structure further enhances the notion that these two resides are involved in interactions with E1. The conserved set of interactions at Gln12/Glu39 suggests that the main chain carbonyl oxygens of Gln12 and Met36 and the conserved water molecules could be also involved in these interactions. Gln12/Glu39 are surrounded by Leu8, Ile15, Met36, Tyr43, Gln57 and Lys68, which are unlikely to contribute into E2/E1 interactions, as these residues are not well conserved in E2 sequences from different papillomaviruses.

The Gln12/Glu39 cluster lies on a side of the N1 domain which is opposite to the side involved in transactivation (and dimerisation), Figure 2c. Notably, the spatial separation of the two functionally important surfaces suggests that E2NT module could be able to interact with E1 at the same time as it interacts through the dimerisation interface with another E2NT module.

The structure reported here for the entire E2 transactivation module, has several implications for understanding of E2 function. It is now possible to map known mutations onto the E2 three-dimensional structure, and to use the knowledge of amino acid conservation and the effects of mutations to assign roles in folding, structure and function to residues. To this end, our results indicate that molecular surfaces involved in transactivation and E1-binding are located at opposite sides of the N1 domain of E2NT, suggesting that both surfaces could be accessed simultaneously by other protein factors. In line with these observations, E1 has been shown to modulate transactivation by directly interacting with E2, leading to repression of transactivation in the presence of excess E1[38]. It is not inconceivable that the docking of E2NT dimer with E1 is sufficient to block further association

The structure shows that the transactivation surface is involved in the formation of the E2NT dimer, which could cross-link E2 molecules bound by their E2CT modules to well-separated DNA sites. Inevitably, such dimerisation would cause DNA to form a loop structure, targeting distally bound transcription factors to regions close to the promoter. While this process has been suggested to be essential for transactivation[36], the definition of interacting surfaces between E2 and other cellular transcription factors requires a great deal of further study.

Our results suggest that the process of DNA loop formation could involve swapping of E2NT modules across E2 dimers bound at separated DNA sites (Fig. 3a-d). The polar components of the monomer-monomer interactions may favour such exchange. Domain swapping is a well-recognised phenomenon that occurs relatively frequently between two individual monomers containing domains connected by a flexible linker [39,40]. E2 is to our knowledge the first example where the swapping event is predicted to occur between dimers.

The dimerisation surface of E2 represents a good target for designing anti-viral drugs, since it is essential for viral transcription, there is no homologous human protein and the residues forming the interface are highly conserved among different viral strains. Dynamic interactions between transcription factors play a central role in the regulation of transcription and replication. Dimerisation, heterodimerisation and the monomer-to-dimer transition may play important roles during the control of the papillomavirus life cycle. These processes themselves can be regulated through phosphorylation, proteolysis, interaction with small ligands or changes in their intracellular concentration. It has been suggested that E2 can regulate the switch between early gene expression and viral genome replication during HPV infection[41]. It is possible that dimerisation of E2NT modules plays an essential role during this process. One scenario would be to activate transcription via induction of DNA loop formation at early stages of the viral life cycle. At later stages, when the concentration of expressed E2 proteins within the cell becomes high and comparable with the K_d for E2 dimer formation, free E2NT modules could compete for

dimerisation with those involved in DNA loop formation and titrate them away, switching off transcription and stimulating replication. It is also possible that other protein factors could be involved in this process, including, for example, E1. Future studies on E2 interactions with viral and cellular transcription factors will reveal the exact structural events involved in transactivation and replication of papillomaviruses. Our results provide an essential step towards understanding these mechanisms.

Materials and Methods

Purification and crystallisation.

Details of the purification and crystallisation of E2NT have been described previously[16]. Briefly, the ORF encoding the N-terminal 201 residues of HPV-16 E2 was cloned into the prokaryotic expression plasmid pET15b downstream of the 20-residue His-tag leader sequence; protein was expressed in *E. coli*BL21(DE3)pLysS and purified using nickel affinity and anion exchange chromatography. Crystals were obtained by hanging drop vapour diffusion with 0.8-1.2M ammonium sulphate, 0.1M triethanolamine pH 8.0-8.3 and 3-5% 2-methyl-2,4-pentanediol. Crystals grew only with very fresh protein preparations and deteriorated in terms of diffraction quality in less than a week. This necessitated freezing and storage of crystals in liquid nitrogen immediately after growth, as discussed above.

Structure determination.

All data were recorded on cryogenically frozen crystals. A native crystal was frozen for which initial data were recorded to 3.4 Å[16]. For the screening of derivatives, crystal stability was even more limiting. Nine crystals were soaked in various heavy atom reagents immediately after growth. The crystals were screened in-house using a MAR research imaging plate on a Rigaku RU200 rotating anode source, by recording

of 15-20% after scaling using SCALEPACK[17] and were stored in liquid nitrogen. The native crystal was transported to EMBL Hamburg where 1.9 Å data were

measured using synchrotron radiation from beam line X11, Table 1. In addition data were recorded at EMBL for the three promising derivatives to about 2.7 Å. Two of these derivatives proved useful in phase determination and the structure was solved by multiple isomorphous replacement with anomalous scattering (MIRAS) at 2.7 Å. The two derivatives were solved independently using the CCP4 suite[43] from the difference Patterson synthesis and by direct methods as implemented in SHELX[44]. Both contained a single heavy atom site. Phases, calculated using MLPHARE, were enhanced by solvent flattening4[5] using a solvent content of 50 %. The resulting high quality density map was easily interpretable and the initial model was built using QUANTA (Molecular Simulations) for all but four residues of the construct, ignoring the His-tag. The model was completed with REFMAC (resolution 20-1.9 Å) using a bulk solvent correction, to an R-factor of 23.3 % (R_{Free} 29.7 % - for 5 % of the data). There are 221 residues in the recombinant protein: the first twenty comprise the His-Tag. The final model contains all but two of the 201 residues of the real protein: residues 125-126 are disordered and lie in a flexible surface loop. Only one residue, His0, of the His-tag has clear density and an ordered conformation. In addition there are 187 water molecules, which were selected using ARP[46] during the course of refinement. The main statistics of the refined model are shown in Table 2.

Analytical ultracentrifugation.

Experiments were carried out in an Optima XL-A ultracentrifuge (Beckman-Coultier, CA, USA) using scanning UV optics. During the experiments, the recombinant E2NT was in 10mM TrisHCl pH 8.0, 5mM DTT, 0.2 mM EDTA, 300 mM NaCl. Data were obtained at rotor speeds of 12,000 and 16,000 rpm, and the time to equilibrium was 10-12 hours. All runs were carried out at 293K, and all radial scans were at a wavelength of 280 nm. Dissociation constants were obtained by nonlinear regression using the Beckman ultracentrifuge software.

References

1. zur Hausen, H. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. *Curr. Top. Microbiol. Immunol.* **186**, 131-156 (1994).

2. Mohr, I. J., Clark, R., Sun, S., Androphy, E. J., MacPherson, P. & Botchan, M. R. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* **250**, 1694-1699 (1990).

3. Giri, I. & Yaniv, M. Structural and mutational analysis of E2 *trans*-activating proteins of papillomaviruses reveals three distinct functional domains. *EMBO J.* **7**, 2823-2829 (1988).

4. McBride, A. & Myers, G. in *Human Papillomaviruses 1997* (eds Myers, G., Sverdrup, F., Baker, C., McBride, A., Münger, K. & Bernard, H.-U.) III-54-III-73 (Theoretical Biology and Biophysics. Los Alamos, 1997).

5. Berg, M. & Stenlund, A. Functional interactions between papillomavirus E1 and E2 proteins. *J. Virol.* **71**, 3853-3863 (1997).

6. Gillette, T. G. & Borowiec, J. A. Distinct roles of two binding sites for the bovine papillomavirus (BPV) E2 transactivator on BPV DNA replication. *J. Virol.* **72**, 5735-5744 (1998).

7. Choo, K. B., Pan, C. C. & Han, S. H. Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: Preferential deletion of the E2 gene and invariable retention of the long control region and E6/E7 open reading frames. *Virology* **161**, 261(1987).

8. Knight, J. D., Li, R. & Botchan, M The activation domain of the bovine papillomavirus E2 protein mediates association of DNA-bound dimers to form DNA loops. *Proc. Natl. Acad. Sci. USA* **88**, 3204-3208 (1991).

9. Li, R., Knight, J. D., Jackson, S. P., Tjian, R. & Botchan, M. R. Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. *Cell* **65**, 493-505 (1991).

Gen. Virol. **77**, 189-198 (1996).

11. E2 sequence database. (1999). http://hpv-web.lanl.gov:

12. Hegde, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505-512 (1992).

13. Hegde, R. S. & Androphy, E. J. Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: Implications for its DNA binding-site selection mechanism. *J. Mol. Biol.* **284**, 1479-1489 (1998).

14. Hegde, R. S., Wang, A. F., Kim, S. S. & Schapira, M. Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein. *J. Mol. Biol.* **276**, 797-808 (1998).

15. Harris, S. F. & Botchan, M. R. Crystal structure of the human papillomavirus type 18 E2 activation domain. *Science* **284**, 1673-1677 (1999).

16. Burns, J. E., Moroz, O. V., Antson, A. A., Sanders, C. M., Wilson, K. S. & Maitland, N. J. Expression, crystallization and preliminary X-ray analysis of the E2 transactivation domain from papillomavirus type 16. *Acta Crystallog.* **D54**, 1471-1474 (1998).

17. Ramakrishan, C. & Ramachandran, G. N. Stereochemical criteria for polypeptide and protein chain conformations. *Biophys. J.* **5**, 909-933 (1995).

18. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291 (1993).

19. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., et al. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542 (1977).

20. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138 (1993).

21. Wiener, M., Freymann, D., Ghosh, P. & Stroud, R. M. Crystal structure of colicin 1a. *Nature* **385**, 461-464 (1997).

22. Frolow, F., Kalb, A. G. & Yariv, J. The structure of a unique, two-fold symmetric, haem-binding protein. *Nature Struct. Biol.* **1**, 453-460 (1994).

23. Mok, Y. K., Gay, G. D., Butler, P. J. & Bycroft, M. Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain. *Protein Sci.* **5**, 310-319 (1996).

24. Foguel, D., Silva, J. L. & de Prat-Gay, G. Characterization of a partially folded monomer of the DNA-binding domain of human papillomavirus E2 protein obtained at high pressure. *J. Biol. Chem.* **273**, 9050-9057 (1998).

25. Gauthier, J.-M., Dostatni, N., Lusky, M. & Yaniv, M. Two DNA-bound E2 dimers are required for strong transcriptional activation and for cooperation with cellular factors in most cells. *The New Biologist* **3**, 498-509 (1991).

26. Estojak, J., Brent, R. & Golemis, E. Correlation of two-hybrid affinity with in vitro measurements. *Mol. Cell. Biol.* **15**, 5820-5829 (1995).

27. Sengchanthalangsy, L., Datta, S., Huang, D., Anderson, E., Braswell, E. & Ghosh, G. Characterisation of the dimer interface of transcription factor NK*k*B p50 homodimer. *Journal of Molecular Biology* **289**, 1029-1040 (1999).

28. Chao, S.-F., Rocque, W. J., Daniel, S., Czyzyk, L. E., Phelps, W. C. & Alexander, K. A. Subunit affinities and stoichiometries of the human papillomavirus type 11 E1:E2:DNA complex. *Biochemistry* **38**, 4586-4594 (1999).

29. Program manual for the Wisconsin package. (8): Madison, Wisconsin, USA: Genetics computer group. (1994).

30. Abroi, A., Kurg, R. & Ustav, M. Transcriptional and replicational activation functions in the bovine papillomavirus type 1 E2 protein are encoded by different structural determinants. *J. Virol.* **70**, 6169-6179 (1996).

31. Cooper, C. S., Upmeyer, S. N. & Winokur, P. L. Identification of single amino acids in the human papillomavirus 11 E2 protein critical for the transactivation or replication functions. *Virology* **241**, 312-322 (1998).

32. Brokaw, J. L., Blanco, M. & McBride, A. A. Amino acids critical for the functions of the bovine papillomavirus type 1 E2 transactivator. *J. Virol.* **70**, 23-29 (1996).

33. Sakai, H., Yasugi, T., Benson, J. D., Dowhanick, J. J. & Howley, P. M. Targeted mutagenesis of the human papillomavirus type 16 E2 transactivation domain reveals separable transcriptional activation and DNA replication functions. *J. Virol.* **70**, 1602-1611 (1996).

34. Breiding, D. E., Sverdrup, F., Grossel, M. J., Moscufo, N., Boonchai, W. & Androphy, E. J. Functional interaction of a novel cellular protein with the papillomavirus E2 transactivation domain. *Mol. Cell. Biol.* **17**, 7208-7219 (1997).

35. Yao, J. M., Breiding, D. E. & Androphy, E. J. Functional interaction of the bovine papillomavirus E2 transactivation domain with TFIIB. *J. Virol.* **72**, 1013-1019 (1998).

36. Semenza, G. L. *Transcription factors and human disease* (Oxford University Press, New York and Oxford, 1998).

37. Ferguson, M. K. & Botchan, M. R. Genetic analysis of the activation domain of bovine papillomavirus protein E2: its role in transcription and replication. *J. Virol.* **70**, 4193-4199 (1996).

38. Le Moal, M. A., Yaniv, M. & Thierry, F. The bovine papillomavirus type 1 (BPV1) replication protein E1 modulates transcriptional activation by interacting with BPV1 E2. *J. Virol.* **68**, 1085-1093 (1994).

39. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. 3D domain swapping: a mechanism for oligomer assembly. *Protein Science* **4**, 2455-2469 (1995).

40. Yang, F., Bewley, C. A., Louis, J. M., et al. Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *Journal of Molecular Biology* **288**, 403-412 (1999).

41. Desaintes, C. & Demeret, C. Control of papillomavirus DNA replication and transcription. *Semin. Cancer. Biol.* **7**, 339-347 (1996).

42. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307-326 (1997).

43. Collaborative Computational Project, N. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallog.* **D50**, 760-763 (1994).

44. Sheldrick, G. M. & Schneider, T. R. SHELXL: high-resolution refinement. *Meth. Enzymol.* **277**, 319-343 (1997).

45. Cowtan, K. D. & Main, P. Phase combination and cross-validation in iterated density-modification calculations. *Acta Crystallog.* **D52**, 43-48 (1996).

46. Lamzin, V. S. & Wilson, K. S. Automated refinement of protein models. *Acta Crystallog.* **D49**, 129-147 (1993).

Table 1

X-ray data and phasing statistics

Data set	Native	UAc	AuCN
Space Group	$P3_121$	$P3_121$	$P3_121$
a ,b (Å)	54.68	54.49	54.58
c (Å)	155.73	155.66	156.50
Resolution (Å)	30-1.9	20-2.7	20 - 2.7
Temperature, K	120	120	120
Wavelength (Å)	0.86	0.86	0.86
Unique reflections	21751	7873	7937
Completeness (%) (outer shell)	98.8 (89.3)	99.8 (96.1)	99.7 (93.8)
R-merge (outer shell)	0.058 (0.339)	0.073 (0.271)	0.061 (0.268)
Phasing Power: (centric / acentric)		1.55 / 2.07	0.95 / 1.40
FOM: MIRAS		0.59	
FOM: DM 20-2.7 Å (2.7 - 1.9 Å)		0.88 (0.61)	
DM: Mean phase change (20-2.7 Å)		32 °	
R-factor (FreeR)	0.223 (0.295)		

Table 2

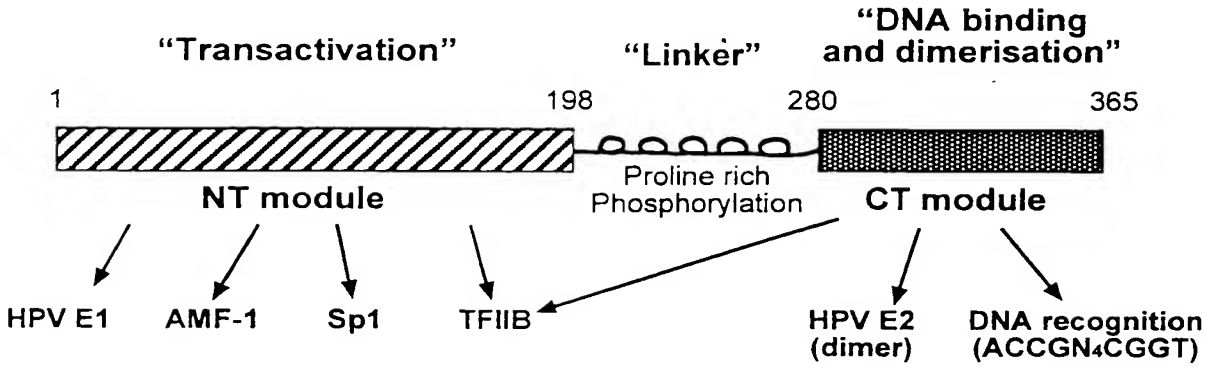
Refinement and model correlation

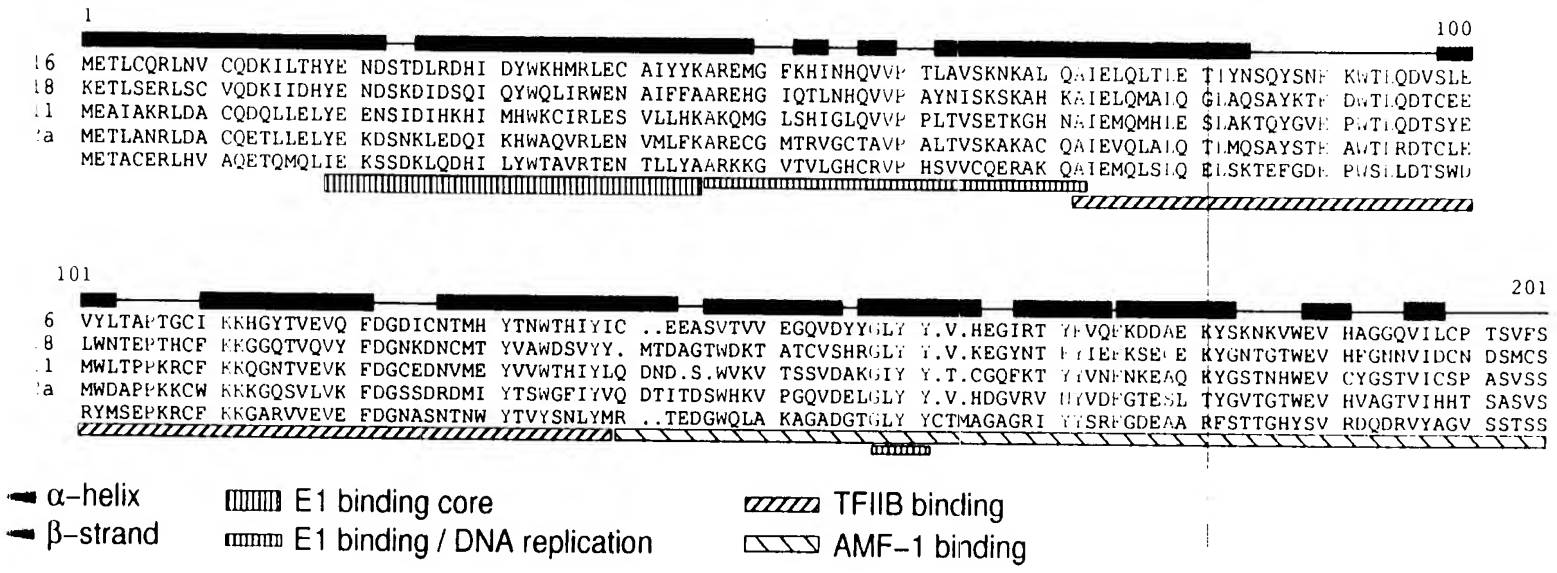
Resolution		1.9 – 10.0 Å
Number of protein atoms		1622
Number of solvent sites		211
Number of reflections used in refinement		20637
Number of reflections used for Rfree calculation		1111
R-factor ‡		0.232
Rfree ‡		0.305
Average atomic B-factor*, Å^2	protein atoms	38.0
	water molecules	48.5
R.m.s. deviations from ideal geometry (Å). Targets in parentheses		
	bond distance	0.013 (0.020)
	angle distance	0.026 (0.040)
	chiral volume	0.142 (0.200)

‡Crystallographic R-factor, $R_{(free)} = \Box \, ||Fo| - |Fc|| \, / \, \Box \, |Fo|$.

Fig 1a

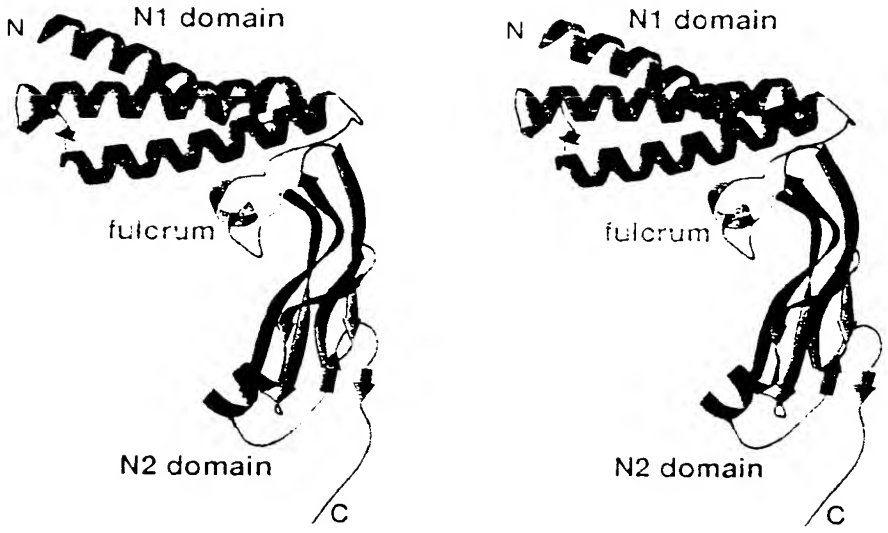
HPV 16 E2 Protein: Functional assignments




1
100
16 METLCQRLNV CQDKILTHYE NDSTDLRDHI DYWKHMRLEC AIYYKAREMG FKHINHQVVP TLAVSKNKAL QAIELQLTLE TIYNSQYSNE KWTLQDVSLE
18 KETLSERLSC VQDKIIDHYE NDSKDIDSQI QYWQLIRWEN AIFFAAREHG IQTLNHQVVP AYNISKSKAH KAIELQMALQ GLAQSAYKTE DWTLQDTCEE
11 MEAIAKRLDA CQDQLLELYE ENSIDIHKHI MHWKCIRLES VLLHKAKQMG LSHIGLQVVP PLTVSETKGH NAIEMQMHLE SLAKTQYGVE PWTLQDTSYE
2a METLANRLDA CQETLLELYE KDSNKLEDQI KHWAQVRLEN VMLFKARECG MTRVGCTAVP ALTVSKAKAC QAIEVQLALQ TLMQSAYSTE AWTLRDTCLE
METACERLHV AQETQMQLIE KSSDKLQDHI LYWTAVRTEN TLLYAARKKG VTVLGHCRVP HSVVCQERAK QAIEMQLSLQ ELSKTEFGDE PWSLLDTSWD
101
201
6 VYLTAPTGCI KKHGYTVEVQ FDGDICNTMH YTNWTHIYIC ..EEASVTVV EGQVDYYGLY Y.V.HEGIRT YFVQFKDDAE KYSKNKVWEV HAGGQVILCP TSVFS
18 LWNTEPTHCF KKGGQTVQVY FDGNKDNCMT YVAWDSVYY. MTDAGTWDKT ATCVSHRGLY Y.V.KEGYNT FYIEFKSEVE KYGNTGTWEV HFGNNVIDCN DSMCS
11 MWLTPPKRCF KKQGNTVEVK FDGCEDNVME YVVWTHIYLQ DND.S.WVKV TSSVDAKGIY Y.T.CGQFKT YYVNFNKEAQ KYGSTNHWEV CYGSTVICSP ASVSS
2a MWDAPPKKCW KKKGQSVLVK FDGSSDRDMI YTSWGFIYVQ DTITDSWHKV PGQVDELGLY Y.V.HDGVRV HYVDFGTESL TYGVTGTWEV HVAGTVIHHT SASVS
RYMSEPKRCF KKGARVVEVE FDGNASNTNW YTVYSNLYMR ..TEDGWQLA KAGADGTGLY YCTMAGAGRI YYSREGDEAA RFSTTGHYSV RDQDRVYAGV SSTSS
α–helix
β–strand
E1 binding core
E1 binding / DNA replication
TFIIB binding
AMF–1 binding


Ala143
Glu142
Ala69
Arg37
Ile73
Gln80
Leu77
Gln76
Glu142
Ala143
Ala143
Glu142
Ile73
Gln76
Leu77
Glu142

4/13


N
N1 domain
fulcrum
N2 domain
C
N
N1 domain
fulcrum
N2 domain
C

5/13

Fig 3a

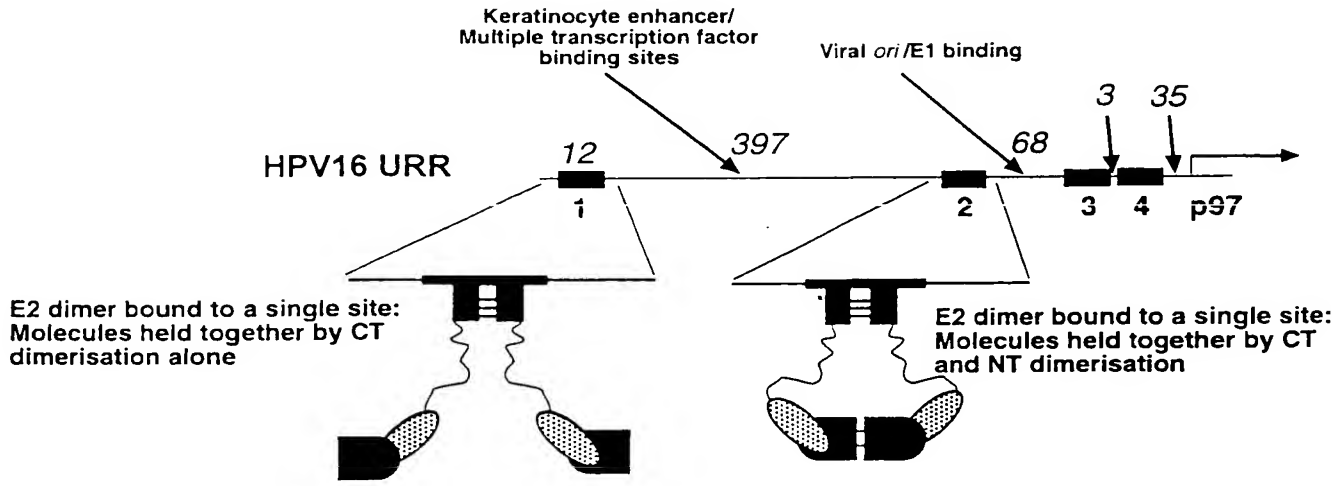
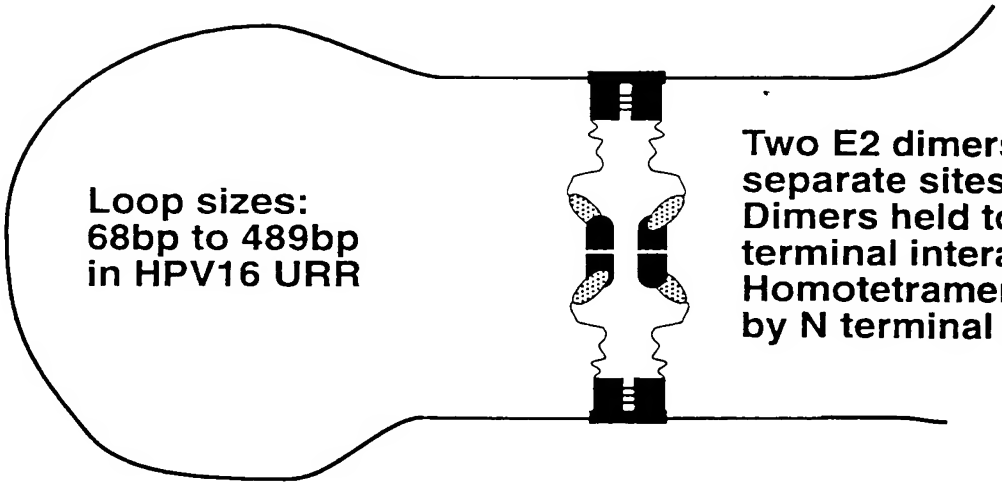

Keratinocyte enhancer/
Multiple transcription factor
binding sites
Viral ori /E1 binding
3
35
HPV16 URR
12
397
68
1
2
3
4
p97
E2 dimer bound to a single site:
Molecules held together by CT
dimerisation alone
E2 dimer bound to a single site:
Molecules held together by CT
and NT dimerisation

Fig 3b


Loop sizes:
68bp to 489bp
in HPV16 URR
Two E2 dimers bound to separate sites:
Dimers held together by C terminal interactions
Homotetramer held together by N terminal dimerisation.

8/13

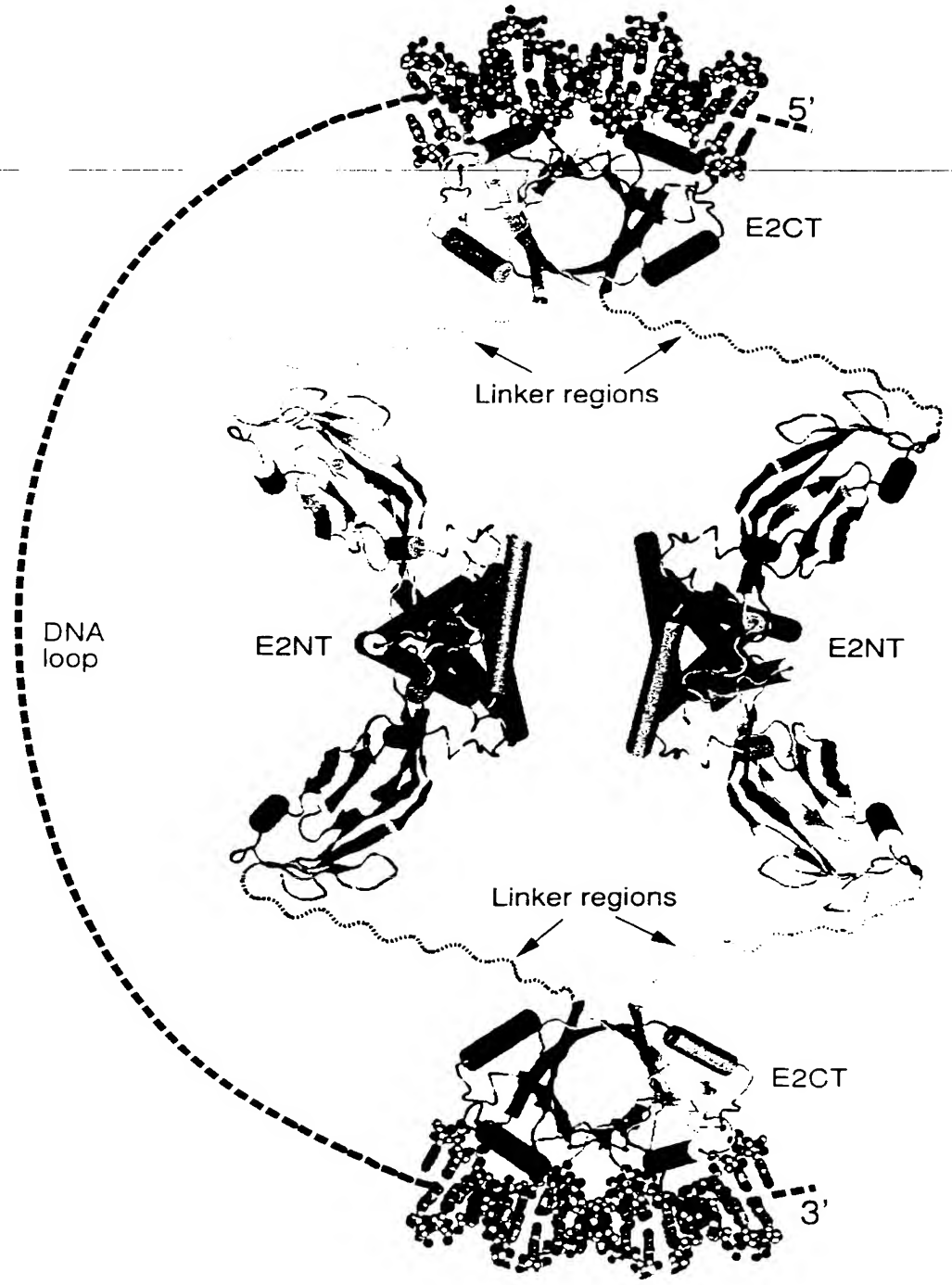
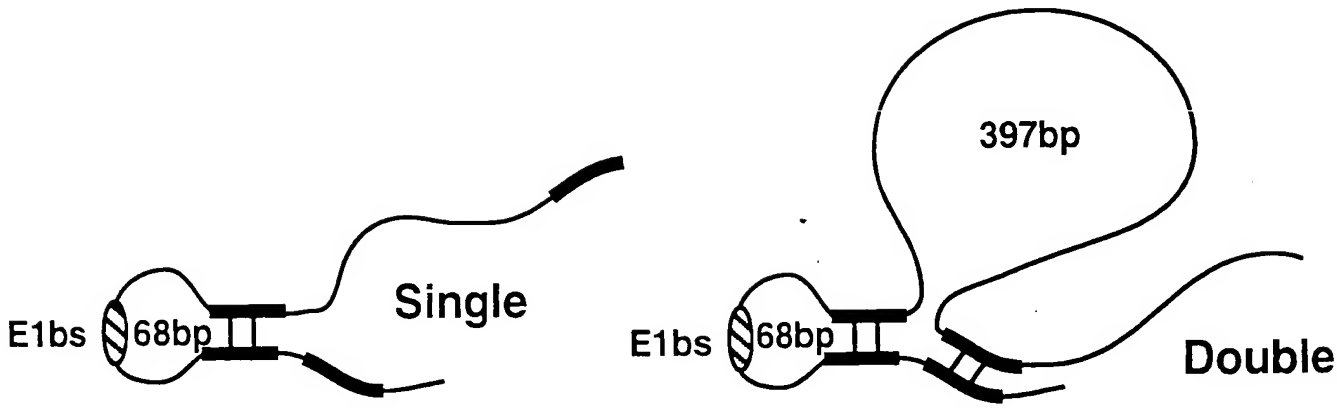
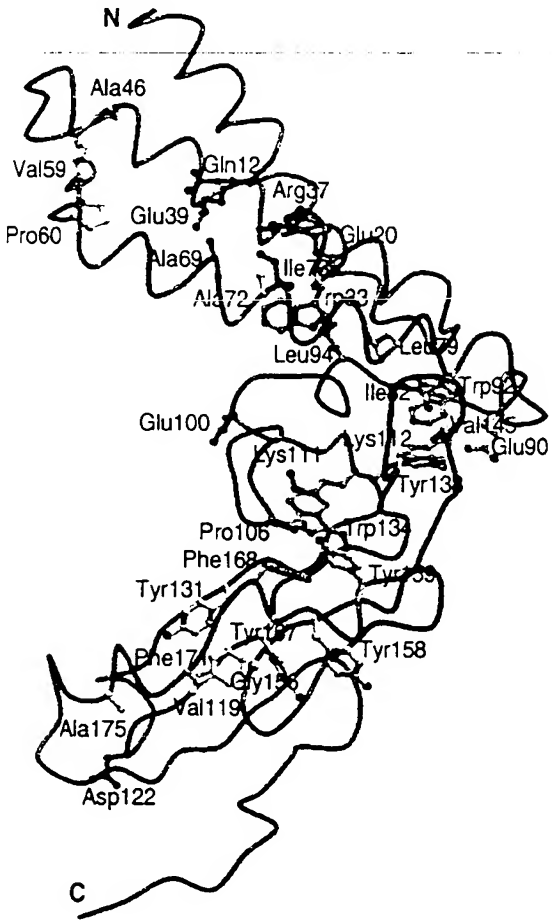
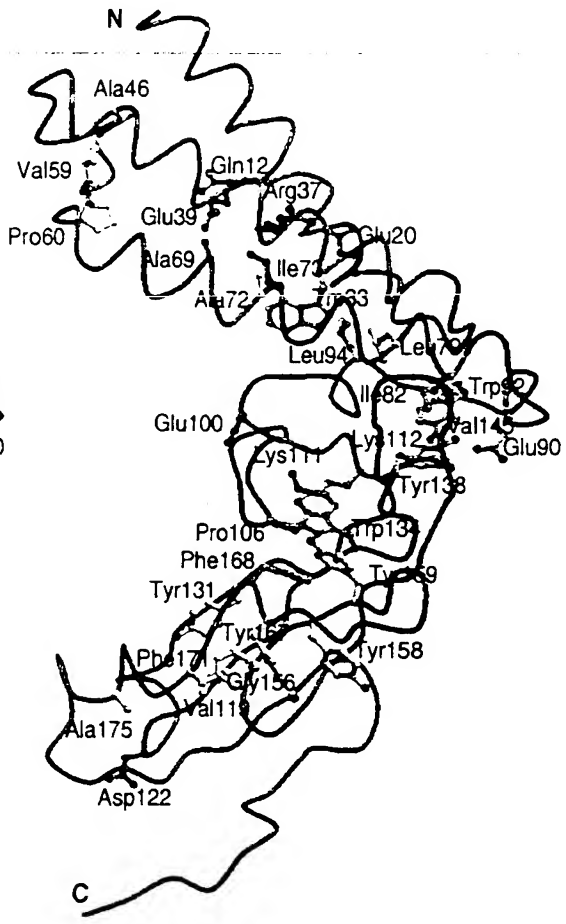
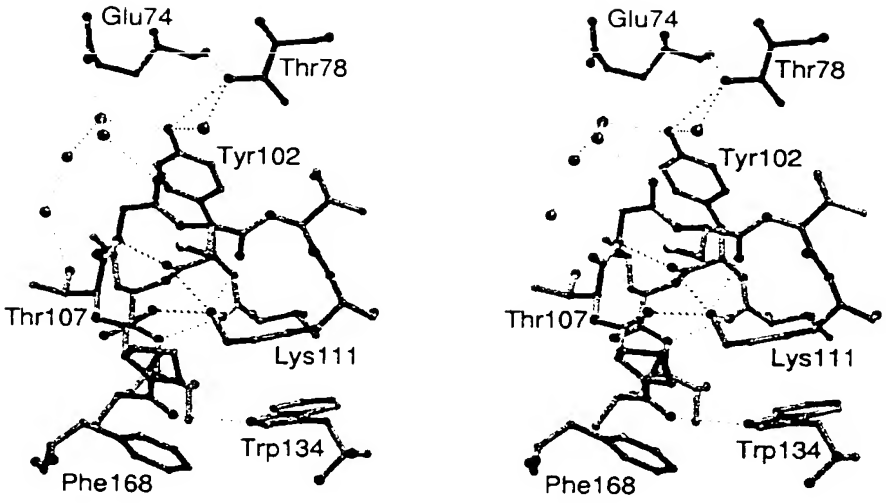

5'
E2CT
Linker regions
DNA
loop
E2NT
E2NT
Linker regions
E2CT
3'

Fig 3a




N
Ala46
Val59
Gln12
Arg37
Pro60
Glu39
Glu20
Ala69
Ile72
Ala72
Trp33
Leu94
Leu79
Ile82
Trp92
Glu100
Val145
Lys112
Glu90
Lys111
Tyr138
Pro106
Trp134
Phe168
Tyr131
Tyr158
Phe171
Gly156
Val119
Ala175
Asp122
C


N
Ala46
Val59
Gln12
Arg37
Pro60
Glu39
Glu20
Ala69
Ile72
Ala72
Trp33
Leu94
Leu79
Ile82
Trp92
Glu100
Val145
Lys112
Glu90
Lys111
Tyr138
Pro106
Trp134
Phe168
Tyr131
Tyr158
Phe171
Gly156
Val119
Ala175
Asp122
C


Glu74
Thr78
Tyr102
Thr107
Lys111
Trp134
Phe168
Glu74
Thr78
Tyr102
Thr107
Lys111
Trp134
Phe168

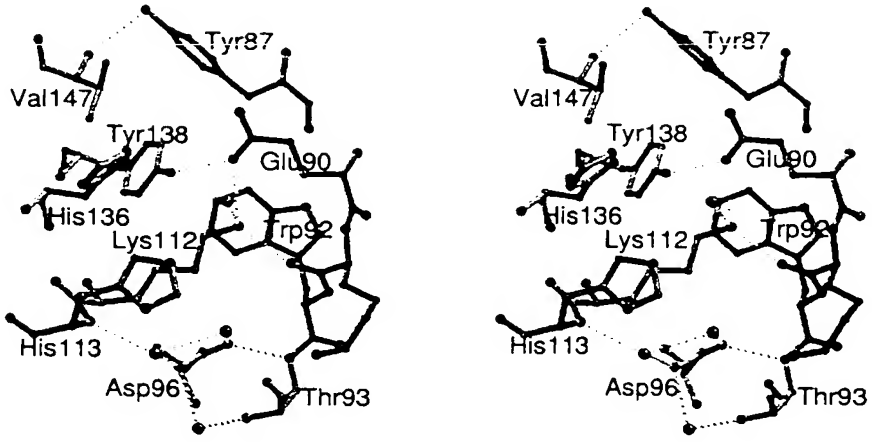
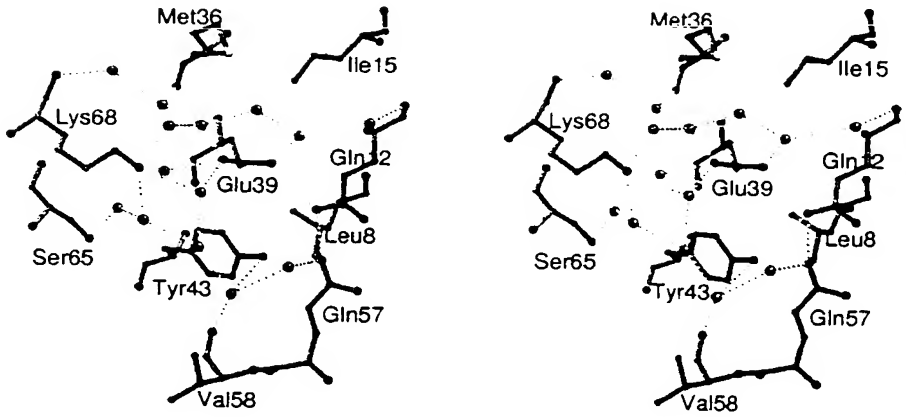

Tyr87
Val147
Tyr138
Glu90
His136
Lys112
Trp92
His113
Asp96
Thr93
Tyr87
Val147
Tyr138
Glu90
His136
Lys112
Trp92
His113
Asp96
Thr93

Fig 4d


Met36
Ile15
Lys68
Glu39
Gln12
Leu8
Ser65
Tyr43
Gln57
Val58
Met36
Ile15
Lys68
Glu39
Gln12
Leu8
Ser65
Tyr43
Gln57
Val58

PCT GB 00 02566

Harrison Goddard Foote

27/9/2000.